
Article:
AlAlaween, W.H., Mahfouf, M. and Salman, A.D. (2016) Predictive modelling of the granulation process using a systems-engineering approach. Powder Technology, 302. pp. 265-274. ISSN 0032-5910

https://doi.org/10.1016/j.powtec.2016.08.049

Predictive Modelling of the Granulation Process Using a Systems-Engineering Approach

Wafa' H. AlAlaween[a*], Mahdi Mahfouf[a] and Agba D. Salman[b]

[a]Department of Automatic Control and Systems Engineering, The University of Sheffield, UK

[b]Department of Chemical and Biological Engineering, The University of Sheffield, UK

(E-mails:* whalalaween1@sheffield.ac.uk; m.mahfouf@sheffield.ac.uk; a.d.salman@sheffield.ac.uk)

(Tel: (+44) (0)1142225611)

Abstract

The granulation process is considered to be a crucial operation in many industrial applications. The modelling of the granulation process is, therefore, an important step towards controlling and optimizing the downstream processes, and ensuring optimal product quality. In this research paper, a new integrated network based on Artificial Intelligence (AI) is proposed to model a high shear granulation (HSG) process. Such a network consists of two phases: in the first phase the inputs and the target outputs are used to train a number of models, where the predicted outputs from this phase and the target are used to train another model in the second phase to lead to the final predicted output. Because of the complex nature of the granulation process, the error residual is exploited further in order to improve the model performance using a Gaussian mixture model (GMM). The overall proposed network successfully predicts the properties of the granules produced by HSG, and outperforms also other modelling frameworks in terms of modelling performance and generalization capability. In addition, the error modelling using the GMM leads to a significant improvement in prediction.

Keywords: High shear granulation; integrated network; radial basis function; ensemble model; Gaussian mixture model.

1. Introduction

Granulation is the enlargement process of fine particles which aims to improve the properties of a powder and facilitate the downstream processes in different industries including chemical, mineral, agriculture, food, and pharmaceutical industries [1]. Research in granulation started more than 5 decades ago, while some of the pioneering research started

much earlier [2]. Although granulation has attracted a lot of interest in academia as well as in industry, it remains an art more than a science where neither the granulation behaviour nor the associated properties can be predicted well in advance, leading, as a result, to inefficient operations and high recycling ratios (waste) [3]. Hence, a deeper understanding of this process either via data, expert knowledge and laws of physics should pave the way for an effective and robust modelling framework to predict the associated process behaviour. Furthermore, exploiting the model in a 'reverse-engineering' framework can lead to 'optimal' control of the process for right-first-time granulation.

Several studies have been devoted to the modelling of the granulation processes, but because of the lack of associated physical equations that should describe the processes, such models are normally either empirical or semi mechanistic models [4]. A regime map has been used to describe the granulation mechanisms, namely; wetting and nucleation, growth and consolidation, and breakage and attrition [5]. Although it was not fully able to represent the associated mechanisms and the properties quantitatively, a comprehensive understanding of the granulation process at the micro level was however reached [6]. Population balance models, by which the rate of change in the number, mass, or volume of the granules during the process are investigated, have also been used to predict the properties of the granules and the granulation process behaviour [7]. Various granules properties and granulation mechanisms have hitherto been investigated. One of the difficulties in performing the population balance based modelling lies in the consideration of all interactions among the granulation mechanisms which are fundamental requirements necessary for shaping the properties of the granules [8]. To characterize such interactions, an integration between a multi-dimensional population balance model and a stochastic method (e.g. Monte Carlo) has been proposed [9]. In addition, the number of the properties that can be monitored using such a technique is limited, with up to three properties only being examined in most published applications. In fact, finding a solution can even prove to be a difficult exercise when more granules properties be included [8].

With the recent advances in computing power, data based modelling approaches have been utilized to model the granulation process, where the main aim is to find a mapping between a set of inputs and outputs instead of deriving the real physical equations [10]. Linear regression models have been employed to predict the properties of granules and to find the optimal set of input parameters [11-12]. Such modelling paradigms are, in fact,

incapable of accounting for the sophisticated nonlinear relationships or even the complex interactions among the inputs parameters that control the granulation process [10]. Artificial Neural Networks (ANN) and Fuzzy Systems have been investigated previously to predict the properties of granules and to scale-up the granulation processes [13-16]. However, because these so-called soft-computing techniques represent powerful interpolators there exist no guarantees that they will perform well beyond the training range [10]. Although these techniques have been extensively employed in various other equally challenging areas (e.g. industrial, academic, and medical) where their effectiveness and efficiency have been demonstrated [17], they have not been well exploited to deal with the challenges and uncertainties in the granulation processes. The reason for this relates to the availability of meaningful data/information needed to derive effective predictive models for granulation. Consequently, these techniques and other data-driven approaches can represent a promising development in dealing with the problems surrounding the granulation process if meaningful information can be extracted from the available data.

In this paper a modelling framework which initially includes the idea of an integrated network is proposed in order to extract meaningful information from a conservative number of granulation data which have been collected from a series of laboratory experiments, where the main motivation behind this framework is to achieve a satisfactory model performance exploiting such a limited amount of real (systematic) data. In order to improve the model performance, the network-based error predictions are characterized using Gaussian mixture model (GMM) to account for any behaviour deemed of a stochastic nature. The paper is organized as follows: a brief description of the experimental work that has been conducted using a high shear granulator is given in Section 2. Section 3 outlines the two modelling stages embedded in the integrated network modelling-based approach including the prediction results. In Section 4, the model is extended to account for previously unmodelled stochastic behaviour via GMM technique. Section 5 draws the overall conclusions relating to the study.

2. Experimental Work

Calcium Carbonate (D_{50}=0.085mm) has been granulated using a high shear Eirich mixer (1 Litre vertical axis granulator with a top-driven impeller, Maschinenfabrik Gustav Eirich GmbH & Co KG, Germany). Polyethylene Glycol (PEG 1000) with a melting point of approximately 40^{o}C has been used as a binder, which has been poured-in in the liquid phase.

The Eirich granulator is equipped with two types of impeller differing in shape. The impeller shape is considered as an input variable in addition to three other variables, as listed in Table 1, while the speed of the vessel is kept constant at 170 rpm. Although, there are many parameters that affect the granulation process, the aforementioned ones are the most crucial for the high shear granulation (HSG) process using specific materials [18]. The levels of each variable have been defined by conducting a set of trial experiments.

Table 1. The inputs and outputs of the granulation process.

Inputs	Inputs' levels	Outputs
Impeller speed	1000, 2000… 6000 (rpm)	Size (µm)
Granulation time	6, 10, and 15 (minutes)	Binder content (%)
L/S ratio (w/w)	13, 14, and 15 (%)	Porosity (%)
Impeller shape	Two different shapes	

The granulation experiments have been carried-out based on a full factorial design of experiments resulting in 108 experiments. After completing the granulation process, the granules have been cooled down at room temperature to solidify the binder. The size of the granules has then been measured. The porosity and the binder content of the granules have been measured using a Pycnometer [19] and the method discussed in [20], respectively. The granulation data are very difficult to model because of (i) the nonlinear behaviour, (ii) the significant effect of the interactions among the input variables, (iii) the uncertainty in the measurements, and (iv) the sparse and limited data. Figure 1 shows examples of data distributions using two variables at a time.

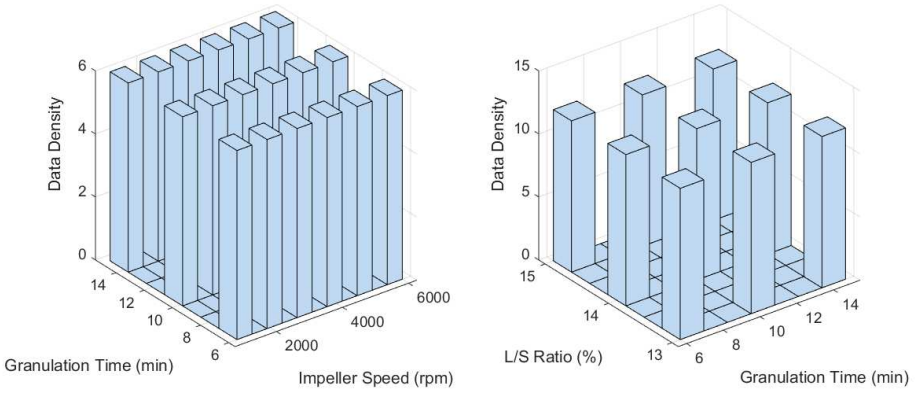


Figure 1. Data density for the granulation time with impeller speed and L/S ratio.

3. Integrated Network

3.1 Integrated Network: Model Development

Recently, the development of computational intelligence has been positively reflected in several disciplines such as medicine and metallurgy [21], where the observed data are utilized to establish data driven models that can replace or complement physical models where they simply do not exist or they may be too complex to elicit. Therefore, the core of such type of modelling rests with process data [22]. In the case of the granulation process, the difficulty stems from the lack of representative information. In addition, the complex input/output relationships may not be captured by the available amount of sparse data [23]. As stated previously, an integrated network as a data-driven model has been proposed in this paper. Developing such a structure does not only consist of mapping the inputs to the outputs, but also discovering knowledge that may not be easy to extract by the already available approaches. The idea of the network relies on having a number of models with different structures, thus, (i) complex input/output relationships could be captured because of the number of functions and weights included [24], (ii) models with different structures could play a complementary role in modelling the possible patterns of the process, and (iii) training the data through two stages could help to extract the associated knowledge required for accurate property predictions [25].

Figure 2 depicts the integrated network architecture for multi-input single-output (MISO). The network relies on predicting the final output using two modelling phases. In phase I, the N inputs (x_n) and the target output (y_T) are used to train M models with different structures. These models can be neural networks or neuro-fuzzy models. The predicted outputs from each model (y_{P1}, y_{P2}... y_{PM}) and the target output (y_T) are then utilized to train another model in phase II to lead to the final predicted output ($\hat{y}_P$), where this model should be capable of modelling linear and non-linear relationships to extract the hidden information and to capture the complex relationships in the original data. The efficiency of the radial basis function (RBF) network has been proved in several application areas [10], thus it has been employed in this paper to model the HSG process.

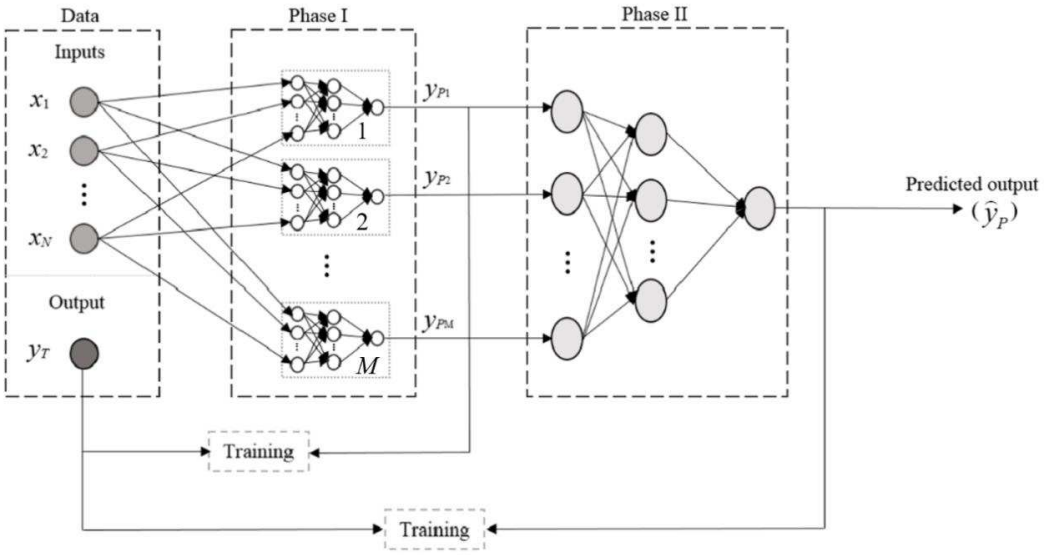


Figure 2. The architecture of the integrated network.

Generally, an RBF network consists of three layers, namely; an input layer, basis functions acting as a hidden layer, and an output layer. Each basis function is a function of the radial distance from a defined centre. These functions are used to map an input vector to its corresponding target. Thus, the predicted output is presented [10] as follows:

$$y_P(x) = \sum_{i=1}^{I} w_i \phi_i(x) + w_0 \tag{1}$$

where w_i and w_0 are the coefficient connecting the i^{th} basis function to the output neuron and the bias, respectively, and ϕ_i is the basis function. A popular selection of such a function is Gaussian [10]. The RBF network has also been used in phase II. Analytically, the two phases of the integrated network are simply a combination of composition and superposition of the basis functions. To prove the capabilities of the presented structure, let us use a single RBF model in phase I, the final predicted output can then be written in the form:

$$\hat{y}_P = \sum_{k=1}^{K} w_k^{(2)} \phi_k \left(\sum_{i=1}^{I} w_i \phi_i(x) + w_0 \right) + w_0^{(2)} \tag{2}$$

the parameters are as defined previously, where the superscript number is used to distinguish the parameters of the second phase from the ones used in the first phase. Assuming that both models are optimized in terms of the number of basis functions and the connecting coefficients, the composite function (2) could minimize the error. It has been proved that the

composite function is dense in a convex data space [26, 27], which means that the difference between the predicted and the target values would be smaller.

Although, the number of models in the first phase has been neglected in the discussion above, it plays a crucial role in the proposed structure. By including the M models, the inner function of (2) could be written as a superposition of the basis function. In a similar way, the theorem that has been presented in [28, 29] demonstrates that the approximated function is also dense in the data space. Thus, the combination of the superposition and composition of the basis functions could considerably improve the model performance.

In this paper, the scaled conjugate gradient (SCG) algorithm has been utilized with the backpropagation network to determine the network parameters for both phases [10, 30]. The root mean square error (RMSE) is employed to select the best network structure (i.e. the number of basis functions) that achieves a trade-off between a good training and generalization capability [30]. These steps are shown in Figure 3.

3.2 Integrated Network: Results and Discussion

A single RBF network is developed here to model the HSG process. The data have been divided into two sets: training and testing. The training data set allows the model to learn the relationships among the granulation inputs and the outputs, while the testing data set assesses its generalization performance. The division of the data into these sets has a significant influence on the performance of the model; by the division of the data one means not only the number of data points in each set but also their distribution in the space under study. Different division methods have hitherto been investigated, including the 10-fold cross-validation technique, but it has been found that dividing the data randomly into a training set (5/6) and a testing set (1/6) was the best methodology in order to develop a meaningful model with a reliable performance [30]. The number of RBFs that has been selected corresponds to the minimum error evaluated via the RMSE. The SCG optimization algorithm has been employed for training. For a single size class (710-1000μm) and using 8 RBFs, the average performance of 10 networks for the binder content is shown in Figure 4, with a RMSE (training, testing) = [0.916, 0.958]. The coefficient of determination value is R^2 (training, testing) = [0.54, 0.31]. These performance measures indicate that the RBF-based model on its own was not able to capture the complex input/output relationships and to achieve adequate generalization capability.

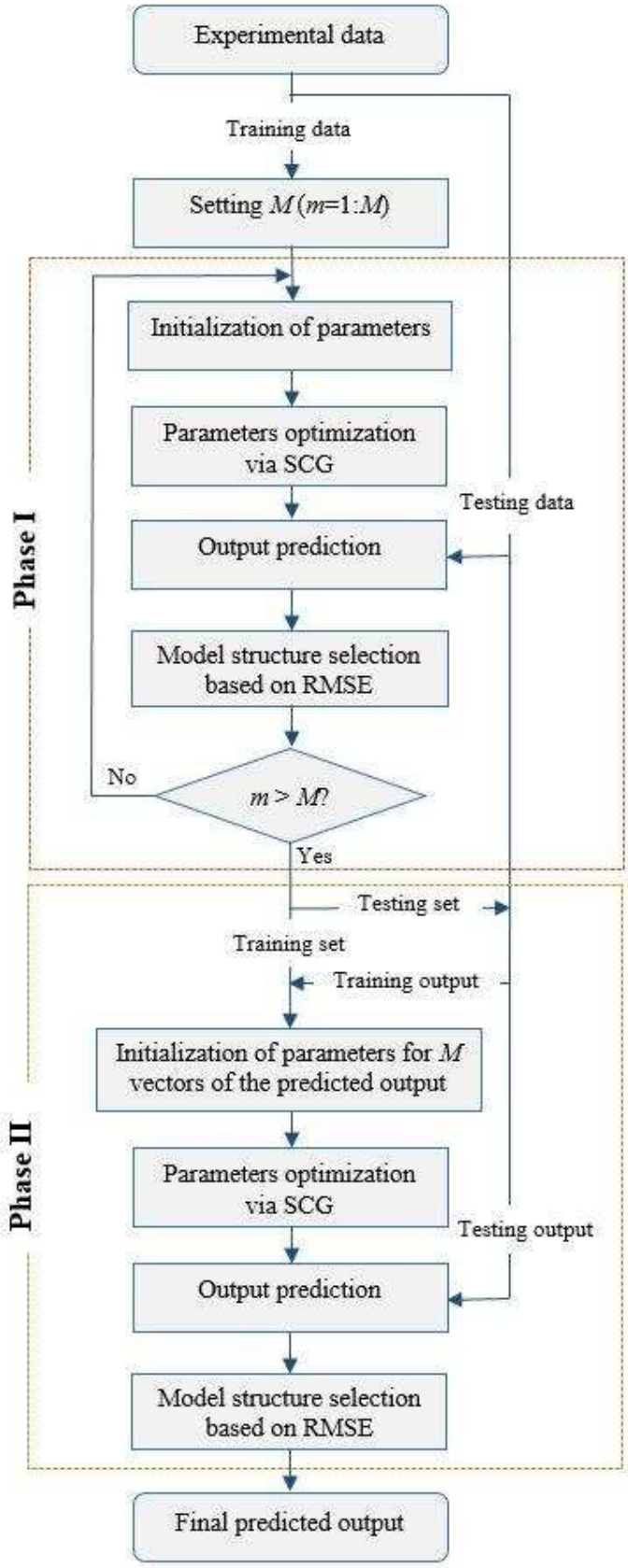


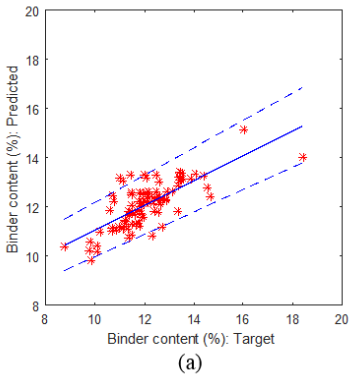
Figure 3. Flow chart of the integrated network.

In a similar manner, the results obtained for the other variables are summarized in Table 2, where the size is represented by its three diameters: D_{10}, D_{50}, and D_{90}.

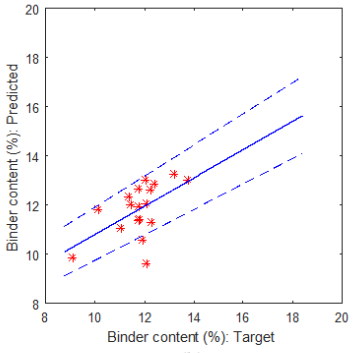
3.2 Integrated Network: Results and Discussion

A single RBF network is developed here to model the HSG process. The data have been divided into two sets: training and testing. The training data set allows the model to learn the relationships among the granulation inputs and the outputs, while the testing data set assesses its generalization performance. The division of the data into these sets has a significant influence on the performance of the model; by the division of the data one means not only the number of data points in each set but also their distribution in the space under study. Different division methods have hitherto been investigated, including the 10-fold cross-validation technique, but it has been found that dividing the data randomly into a training set (5/6) and a testing set (1/6) was the best methodology in order to develop a meaningful model with a reliable performance [30]. The number of RBFs that has been selected corresponds to the minimum error evaluated via the RMSE. The SCG optimization algorithm has been employed for training. For a single size class (710-1000μm) and using 8 RBFs, the average performance of 10 networks for the binder content is shown in Figure 4, with a RMSE (training, testing) = [0.916, 0.958]. The coefficient of determination value is R^2 (training, testing) = [0.54, 0.31]. These performance measures indicate that the RBF-based model on its own was not able to capture the complex input/output relationships and to achieve adequate generalization capability. In a similar manner, the results obtained for the other variables are summarized in Table 2, where the size is represented by its three diameters: D_{10}, D_{50}, and D_{90}.

To improve the prediction performance of the RBF network, an ensemble model has been implemented [24], where the outputs of multiple networks are combined, commonly, by a simple averaging method [31]. Ten RBF networks have been initialized using different number of basis functions, as listed in Table 2, and different values for the connecting coefficients. The prediction performance for the ensemble model is superior to that of the single network above, with an improvement of 28% in RMSE for the testing data set.



(a)



(b)

Figure 4. The RBF model for the binder content: (a) Training, (b) testing (with 10% bands).

Furthermore, the new integrated network structure based on the 10 RBF models in the first phase, having the same structure as that of the ensemble model, and a single RBF model in the second phase has been established. The integrated network performance for the binder content (710-1000µm) is R^2 (training, testing) = [0.75, 0.74], as shown in Figure 5, while examples of the predicted and the experimental distributions for all the investigated properties are presented in Figure 6. The obtained results prove the ability of the integrated network in dealing with the difficulties and complexity of the granulation behaviour. The R^2 value for the integrated network is approximately twice the value for the single RBF network, whereas the overall improvement over the ensemble model is approximately 34%.

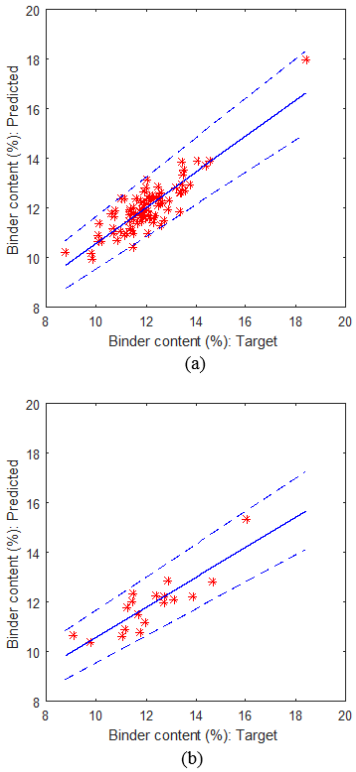


Figure 5. The integrated network based on 10 RBF models for the binder content: (a) Training, (b) testing (with 10% bands).

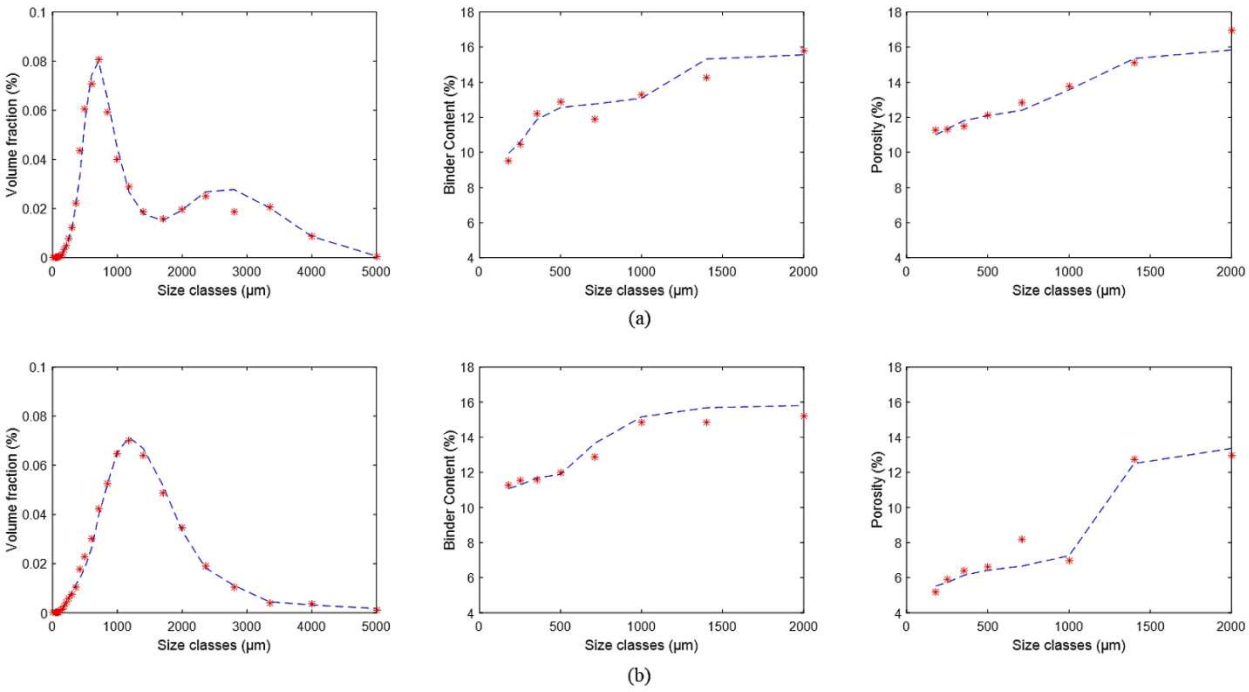


Figure 6. The integrated network: the predicted (*) and the experimental (--) distributions for the size, binder content and porosity (a) using impeller type II, speed=2000rpm, L/S ratio (w/w)=14% and granulation time=10min, (b) using impeller type II, speed=6000rpm, L/S ratio (w/w)=15% and granulation time=15min.

4. Error Modelling Using Gaussian Mixture Model

4.1 Error Modelling Using Gaussian Mixture Model: Model Development

Occasionally, the error can play a significant role in refining the model by eliciting the information that may be hidden because of the implicit assumption that the error is normally distributed. Different error models have already been proposed previously [32-34]. The model that depends on the GMM has been demonstrated to be an efficient model in terms of error characterization. Due to the inherent complexity of the granulation process with highly nonlinear behaviour of the process and measurement uncertainties, the GMM has been selected to provide a deeper insight into the probability density function. Moreover, it has the ability to approximate any probability density function with a reasonable accuracy using a sufficient number of Gaussian components, which can lead to the optimal model refinement [30, 34]. Figure 7 presents a schematic representation of the incorporation of the integrated network and the error characterization using GMM.

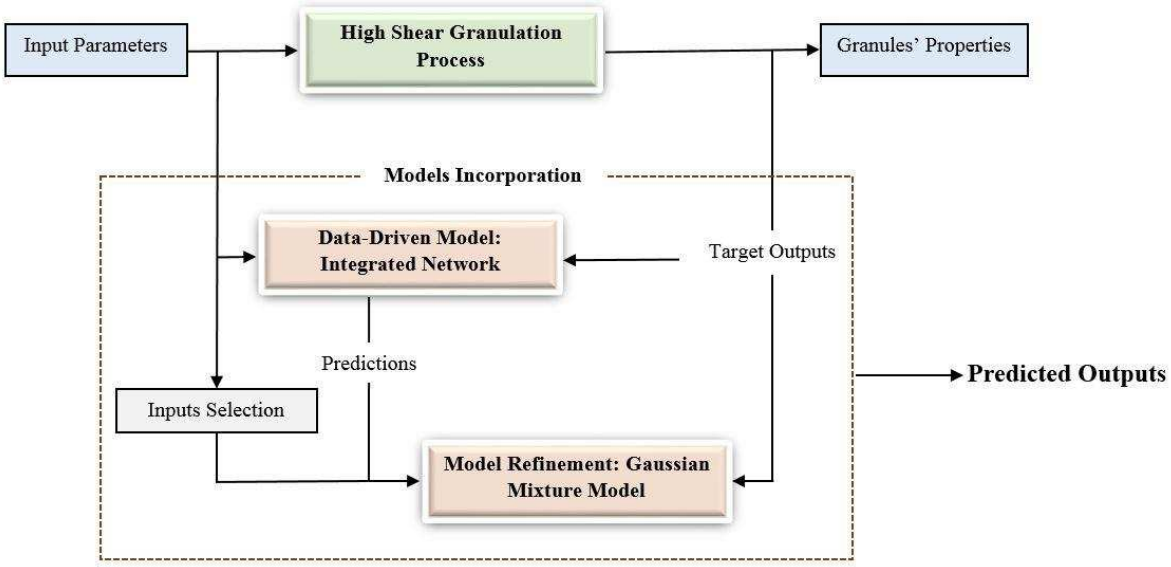


Figure 7. The incorporation of the integrated network and the error characterization framework.

The GMM, in general, is a stochastic model that can be represented as a linear combination of Gaussian components, where each component has its own mean and covariance. For a predefined number of Gaussian components (J), the GMM is simply presented as follows [35]:

$$p\left(x^{e}\right)=\sum_{j=1}^{J}\pi_{j}N\left(x^{e}\setminus\mu_{j}^{e},\Sigma_{j}^{e}\right) \tag{3}$$

where x^e is the error data which contain the selected inputs and the error vector. The number of inputs that will be included in the error characterisation should be small [34], since the main effect of the inputs has been considered in the integrated network. μ_j^e, Σ_j^e, π_j are the mean, the covariance, and the mixing coefficient of the j^{th} Gaussian component, respectively. The superscript e is used to distinguish the parameters that are defined in the error model. To define the optimal values of these parameters, the log likelihood function should be maximized [30]. Therefore, the optimal parameters are given by the following set of equations:

$$\xi(z_{dj})=\frac{\pi_{j}N\left(x_{d}^{e}\mid\mu_{j}^{e},\Sigma_{j}^{e}\right)}{\sum_{j=1}^{J}\pi_{d}N\left(x_{d}^{e}\mid\mu_{j}^{e},\Sigma_{j}^{e}\right)}\quad,\qquad\forall j$$

$$\left\{\begin{array}{l}\mu_{j}^{e}=\dfrac{\sum_{d=1}^{D}\xi(z_{dj})x_{d}^{e}}{\sum_{d=1}^{D}\xi(z_{dj})}\\[2ex]\Sigma_{j}^{e}=\dfrac{\sum_{d=1}^{D}\xi(z_{dj})\left(x_{d}^{e}-\mu_{j}^{e}\right)\left(x_{d}^{e}-\mu_{j}^{e}\right)^{T}}{\sum_{d=1}^{D}\xi(z_{dj})}\\[2ex]\pi_{j}=\dfrac{\sum_{d=1}^{D}\xi(z_{dj})}{D}\end{array}\right\},\qquad\forall j \tag{4}$$

where $\xi(z_{dj})$ is the probability that the d^{th} data point belongs to the j^{th} Gaussian component, and z_{dj} is a J-dimensional latent variable, which is equal to 1 when the d^{th} data point is covered by the j^{th} component where the other elements are zero. Deriving the analytical solution for these equations is a rather 'tricky' exercise but suffice to say that one of the most common methods for finding a solution for such a set of equations is the Expectation Maximization (EM) algorithm [36]. Starting by carefully initializing the parameters using K-means clustering, $\xi(z_{dj})$ value can be estimated using the initialized parameters, the so called

E-step. Accordingly in the M-step, the $\xi(z_{dj})$ value is utilized to re-evaluate the parameters. The revised parameters are then utilized to update the $\xi(z_{dj})$ value. Such a procedure is reiterated until the algorithm converges, or alternatively the maximum number of iteration is reached [36].

However, the number of Gaussian components (J) has to be defined. Thus, the EM algorithm is utilized to select the number of components. Bayesian information criterion (BIC) has been adopted in this paper as a performance criterion for selecting the best number of components [37]. Finally, the conditional error mean, which is an indication of the bias and its value, and the conditional standard deviation are calculated using numerical methods [34]. Generally, these methods are considered to be computationally taxing, however, it seems not to be the case in this paper, particularly with a small data set [38]. Adding the conditional mean to the predicted output is a compensation for the bias, which can improve the prediction performance, whereas the conditional standard deviation is used to set the confidence level [34]. Figure 8 summarizes the steps of the error characterization model.

4.2 Error Modelling Using Gaussian Mixture Model: Results and Discussion

In order to improve the presented models performance by characterising the error employing the GMM, two granulation input variables out of a total of four are included. The combination that gives the maximum error compensation (i.e. the minimum RMSE) is finally chosen. Following the steps that were summarized in Figure 8, the impeller speed and the granulation time are utilized in addition to the error vector that results from the integrated network to develop the error model for the binder content. The selection of these parameters by the optimisation process was expected, since the effects of these parameters appear to be incompatible as reported by previous research [39-40]. This may perhaps relate to the interaction among the parameters which may result in the unpredictable behaviour of the granulation process. The training data are employed to train the GMM whereas the testing data are kept hidden. The best number of Gaussian components for the binder content is 10. Figure 9 shows the prediction results after bias compensation for the binder content (710-1000μm) with a 95% confidence interval.

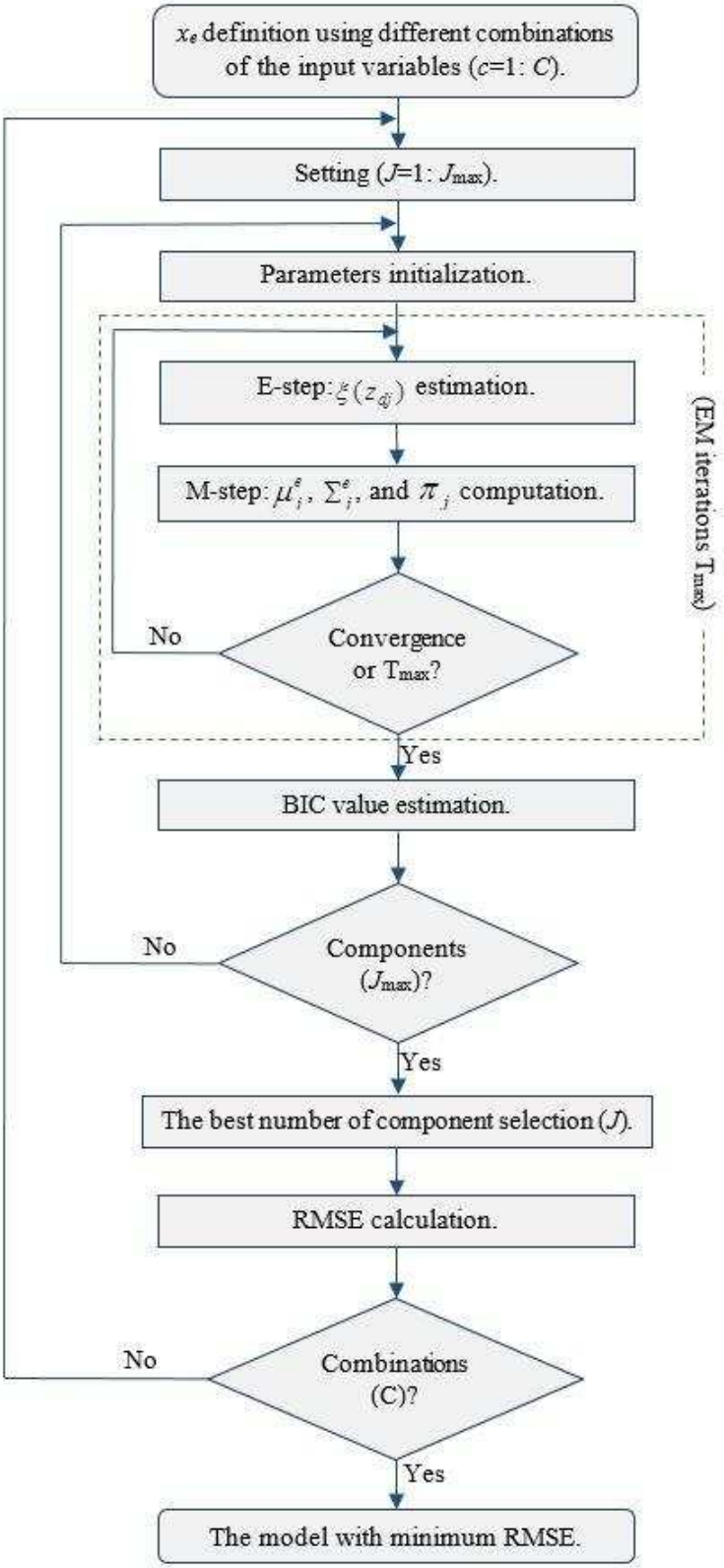


Figure 8. Flowchart of the error characterisation model.

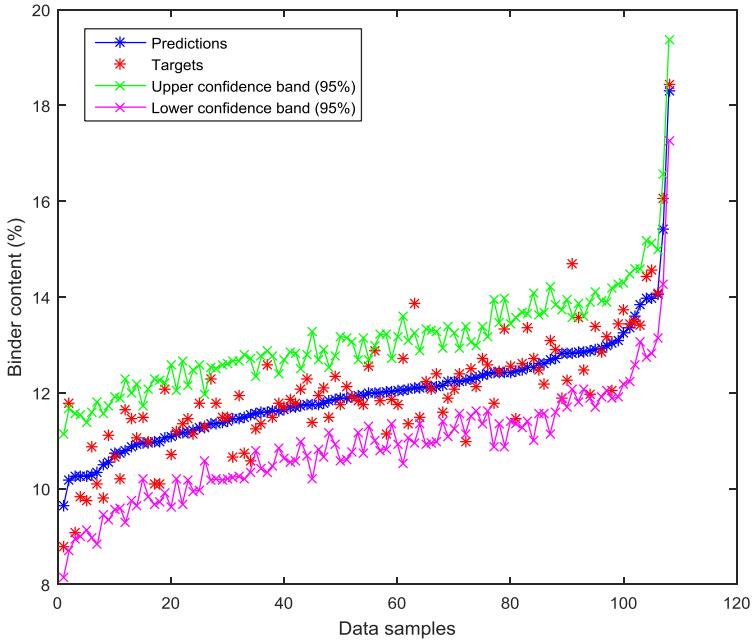


Figure 9. The prediction performance using the integrated network for the binder content after bias compensation (with a 95% confidence interval).

The output predictions with bias compensation presented in Figure 9 elucidates a satisfactory performance, where most of the predictions (96%) are laying within the 95% confidence interval. The overall improvement that is gained by employing the GMM is of approximately 14% in RMSE which is due to the ability of the GMM to capture the inherent undetected stochastic behaviour of the granulation process. The GMM for error characterization has also been adopted to improve the performance of the single RBF and the ensemble model, resulting in a significant improvement for each model, as summarized in Table 2. However, it is evident that the proposed integrated network outperforms these models, even without bias compensation.

Table 2 lists the results of the models for all the investigated outputs. It shows that the performances for the binder content and porosity are generally worse than the ones for the size represented by three diameters (D_{10}, D_{50}, D_{90}). The heterogeneity of the granules from the same batch but different size classes has been demonstrated in the previous research [41-42]. It has also been shown that the same size granules have different values for these properties, such differences may be due to the uncertainties in the measurements, the heterogeneity of the same size granules or both [43]. Such uncertainties and heterogeneity may be the reasons behind the low prediction performances for these properties. Further

investigations have to be performed to clarify this issue and to consider it in the developed model.

Table 2. The performances of the models represented by RMSE and R^2.

Models	Output	Binder content (710-1000µm)		Porosity (710-1000µm)		Size (mm)					
						D_{10}		D_{50}		D_{90}	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
RBF	R^2	0.54	0.31	0.44	0.27	0.53	0.31	0.64	0.33	0.58	0.41
	RMSE	0.92	0.96	1.83	2.82	0.15	0.21	0.26	0.72	1.12	1.01
	No. BFs	8		4		6		9		8	
RBF with bias compensation	R^2	0.64	0.55	0.54	0.15	0.68	0.34	0.72	0.34	0.65	0.37
	RMSE	0.81	0.82	1.67	3.05	0.12	0.19	0.23	0.72	1.03	1.07
	No. GCs	9		8		5		8		9	
Ensemble	R^2	0.59	0.45	0.63	0.43	0.73	0.72	0.77	0.77	0.84	0.78
	RMSE	0.99	0.67	1.71	1.82	0.12	0.08	0.31	0.2	0.81	0.73
	No. BFs	(10, 5, 1, 4, 1, 3, 7, 9, 5 and 6)		(6, 4, 5, 15, 11, 7, 13, 4, 3 and 6)		(4, 11, 7, 4, 5, 10, 9, 13, 3 and 3)		(15, 13, 10, 15, 7, 6, 3, 4, 7 and 11)		(12, 11, 14, 6, 9, 15, 4, 2, 9 and 3)	
Ensemble with bias compensation	R^2	0.63	0.61	0.67	0.29	0.79	0.62	0.76	0.8	0.86	0.79
	RMSE	0.87	0.57	1.57	1.99	0.11	0.11	0.28	0.18	0.68	0.69
	No. GCs	9		8		9		9		5	
Integrated network	R^2	0.75	0.74	0.74	0.74	0.86	0.9	0.83	0.87	0.92	0.89
	RMSE	0.62	0.9	1.31	1.91	0.08	0.04	0.23	0.14	0.45	0.67
	No. BFs (2)	10		8		6		6		8	
Integrated network with bias compensation	R^2	0.82	0.74	0.76	0.74	0.87	0.92	0.86	0.84	0.93	0.89
	RMSE	0.52	0.86	1.26	1.86	0.08	0.04	0.21	0.18	0.41	0.66
	No. GCs	10		6		6		7		8	

1. 'No. BFs' stands for the number of basis functions.
2. 'No. GCs' stands for the number of Gaussian Components.
3. 'No. BFs (2)' represents the number of basis function in the second phase of the integrated network where the 10 models in the first phase have the same structure as the ones in the ensemble model.

To prove the effectiveness and efficiency of the proposed integrated network in dealing with the challenges and difficulties surrounding the granulation process, the network has been used to predict the outputs for new granulation data. Thus, 10 new experiments have been conducted using different input settings, but within the examined ranges. The prediction outputs from the integrated network with error correction have been compared with the measured ones. For the binder content (710-1000µm), Figure 10 shows the performance of the model for the validation data, where the R^2 (=0.76) is comparable to the one for the testing set for the same property. Most of the predictions for the validation data fit properly within a 95% confidence interval. Similarly for all the outputs, the performance for the validation data is close to the one of the testing set. Figure 11 shows an example of the

predicted and the experimental distributions for all the investigated properties for one of the new experiments, where the proposed model has successfully predicted the properties.

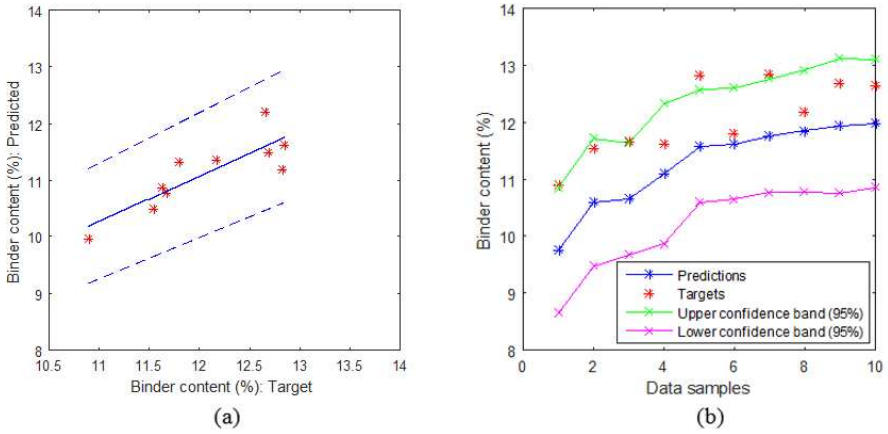


Figure 10. The performance for the binder content for the validation data using the integrated network: (a) predicted versus target, (b) with 95% confidence interval.

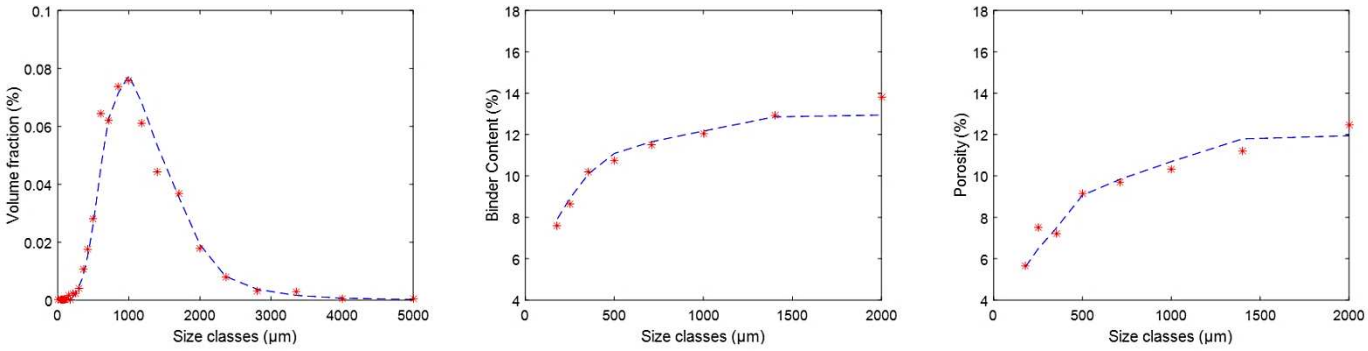


Figure 11. The proposed framework: the predicted (*) and the experimental (--) distributions for the size, binder content and porosity (using impeller type I, speed=4400rpm, L/S ratio (w/w)=13.6% and granulation time=12min).

5. Conclusions

Modelling the granulation process is not a trivial task because of the complex nature of such a process and the lack of physical representation of its behaviour, where the modelling approaches of the granulation process that have received the most attention have hitherto focused on analytical and numerical based techniques in the form of empirical and semi-mechanistic models. Moreover, the limited and sparse amount of data is considered as

another difficulty in modelling the granulation process using data-driven models, especially for some industrial applications, including the pharmaceutical industry, where the acquiring of such data alone can be an expensive enterprise. In this research work, a new integrated modelling framework has been developed to predict the properties of the granules produced by HSG. The integrated network predicts the outputs by modelling and training the data in two consecutive phases. Such a structure is able to extract relevant information from a conservative number of data points, it also has the ability to capture the complex input/output relationships in the original data because of the number of basis functions and weights involved. Moreover, one of the major obstacles for developing data driven models; defining the best structure, has been overcome by using different modelling structures in the first phase of the network. The efficiency of the new network has been demonstrated and validated by accurately predicting the properties of the granules (size, binder content, and porosity). Characterizing the resulted error using GMM has then been integrated in the original model structure to deal with any potential bias in the predicted outputs. Such a characterization has the ability to reveal the stochastic behaviour which has been utilised for further model refinement. It has been shown that most of the output predictions for all the properties fit adequately within a 95% confidence interval.

The framework efficiency, which in our view emanates from the integration of deterministic and stochastic modelling, has been successfully demonstrated in this study. When compared to the computationally expensive models that have been mentioned previously, for example population balance models, this modelling framework has accurately predicted the properties of the granules within a reasonable time. However, there is a strong demand for improving the interpretability of the granulation process and for dealing with uncertainties. In the future, it would also be worth incorporating the proposed network with other models such as neuro-fuzzy models, where the system is described linguistically in a transparent way that can easily be understood by users and therefore ‘owned’. In addition, these models are capable of dealing with uncertainties more effectively; by uncertainty here one means not only uncertainties in the measurements but also uncertainties which result from the heterogonous distribution of the binder content and porosity during the granulation process. Moreover, the integration between the presented network and other physical based models (e.g. population balance models and computational fluid dynamics) will be advantageous, particularly for the ill-defined granulation process, where the integrated network can circumvent the limitations of such models (e.g. the number of outputs and the

execution time). In contrast to data-driven models, physical model can compensate for the limited number of data points. Furthermore, incorporating these models would be very beneficial for scaling-up processes.

In summary, the presented framework represents a promising development in the granulation process, as it lays the foundation not only for accurate predictions but also for the optimal control of the process and the industries, where such a process is crucial to determining the quality of the downstream product. Exploiting the model in a 'reverse-engineering' framework would not only lead to process control but also to optimising product quality.

Acknowledgment

The Authors would like to thank Maschinenfabrik Gustav Eirich GmbH & Co KG (Germany) for providing the Eirich mixer that has been used for the experimental work.

References

[1] G. Reynolds, J. Fu, Y. Cheong, M. Hounslow, A. Salman, Breakage in granulation: A review, Chemical Engineering Science 60 (2005) 3969-3992.

[2] C. Capes, P. Danckwerts, Granule formation by the agglomeration of damp powders: Part 1. The mechanism of granule growth, Transaction Institute of Chemical Engineering 43 (1965) 116-124.

[3] G. Walker, Future Development in Drum Granulation Modelling. In: A. Salman, M. Hounslow, J. Seville (Eds), Granulation, Oxford: Elsevier, 2007, pp. 249-254.

[4] I. Bjorn, A. Jansson, M. Karlsson, S. Folestad, A. Rasmuson, Empirical to mechanistic modelling in high shear granulation, Chemical Engineering Science 60 (2005) 3795-3803.

[5] S. Iveson and J. Litster, Growth regime map for liquid-bound granules, American Institution of Chemical Engineers Journal 44 (1998) 1510-1518.

[6] W. Tu, A. Ingram, J. Seville, S. Hsiau, Exploring the regime map for high-shear mixer granulation, Chemical Engineering Journal 145 (2009) 505-513.

[7] C. Sanders, A. Willemse, A. Salman, M. Hounslow, Development of a predictive agglomeration model, Powder Technology 138 (2003) 18-24.

[8] S. Iveson, Limitations of one-dimensional population balance models of wet granulation processes, Powder Technology 124 (2002) 219-229.

[9] A. Braumann, M. Goodson, M. Kraft, P. Mort, Modelling and validation of granulation with heterogeneous binder dispersion and chemical reaction, Chemical Engineering Science 62 (2007) 4717- 4728.

[10] C. Bishop, Neural Networks for Pattern Recognition, Oxford: Clarendon Press, New York, 1995.

[11] J. Westerhuis, P. Coenegracht, C. Lerk, Multivariate modelling of the tablet manufacturing process with wet granulation for tablet optimization and in-process control, International Journal of Pharmaceutics 156 (1997) 109-117.

[12] Y. Miyamoto, S. Ogawa, M. Miyajima, M. Matsui, H. Sato, K. Takayama, T. Nagai, An application of the computer optimization technique to wet granulation process involving explosive growth of particles, International Journal of Pharmaceutics 149 (1997) 25-36.

[13] E. Murtoniemi, J. Yliruusi, P. Kinnunen, P. Merkku, K. Leiviska, The advantages by the use of neural networks in modelling the fluidized bed granulation process, International Journal of Pharmaceutics 108 (1994) 155-164.

[14] S. Watano, Y. Sato, K. Miyanami, Application of a neural network to granulation scale-up, Powder Technology 90 (1997) 153-159.

[15] R. Mansa, R. Bridson, R. Greenwood, H. Barker, J. Seville, Using intelligent software to predict the effects of formulation and processing parameters on roller compaction, Powder Technology 181 (2008) 217-225.

[16] H. Yu, J. Fu, L. Dang, Y. Cheong, H. Tan, H. Wei, Prediction of the Particle Size Distribution Parameters in a High Shear Granulation Process Using a Key Parameter Definition Combined Artificial Neural Network Model, Industrial and Engineering Chemistry Research, 54 (2015) 10825-10834.

[17] H.J. Kim, M. Mahfouf, Y.Y. Yang, Modelling of hot strip rolling process using a hybrid neural network approach, Journal of Materials Processing Technology 201 (2008) 101-105.

[18] T. Chitu, D. Oulahna, M. Hemati, Wet granulation in laboratory scale high shear mixers: effect of binder properties, Powder Technology 206 (2011) 25–33.

[19] J. Fu, Y. Cheong, G. Reynolds, M. Adams, A. Salman, M. Hounslow, An experimental study of the variability in the properties and quality of wet granules, Powder Technology 140 (2004) 209-216.

[20] P.C. Knight, T. Instone, J.M.K. Pearson, M.J. Hounslow, An investigation into the kinetics of liquid distribution and growth in high shear mixer agglomeration, Powder Technology 97 (1998) 246-257.

[21] C.S. Nunes, M. Mahfouf, D.A. Linkens, J.E. Peacock, Modelling and multivariable control in anaesthesia using neural-fuzzy paradigms Part I. Classification of depth of

anaesthesia and development of a patient model, Artificial Intelligence in Medicine 35 (2005) 195-206.

[22] Y.Y. Yang, M. Mahfouf, Q. Zhang, Optimal input selection for neural fuzzy modelling with application to Charpy energy prediction, IEEE International Conference on Fuzzy Systems (2011) 27-30, Taiwan.

[23] Y.Y. Yang, D.A. Linkens, M. Mahfouf, A.J. Rose, Grain growth modelling for continuous reheating process–a neural network-based approach, ISIJ International 43 (2003) 1040-1049.

[24] D. Opitz, R. Maclin, Popular ensemble methods: An empirical study, Journal of Artificial Intelligence Research **11** (1999) 169-198.

[25] S. Gaffour, M. Mahfouf, Y.Y. Yang, 'Symbiotic' data-driven modelling for the accurate prediction of mechanical properties of alloy steels, 5th IEEE International Conference Intelligent Systems (2010) 31-36.

[26] M. Mateljevic, M. Pavlovic, The best approximation and composition with inner function, Michigan Mathematical Journal 42 (1995) 367-378.

[27] G. Cybenko, Continuous valued neural networks with two hidden layers are sufficient, Technical Report, Department of Computer Science, Tufts University, 1988.

[28] G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals, and Systems 2 (1989) 303-314.

[29] H. N. Mhaskar, C.A. Micchelli, Approximation by superposition of sigmoidal and radial basis functions, Advances in Applied Mathematics 13 (1992) 350-373.

[30] C. Bishop, Pattern recognition and machine learning, Springer, New York, 2006.

[31] Z. Zhao, Y. Zhang, Design of ensemble neural network using entropy theory, Advances in Engineering Software 42 (2011) 838-845.

[32] J. A. Mauricio, Computing and using residuals in time series models, Computational Statistics & Data Analysis 52 (2008) 1746-1763.

[33] J. Oliveira, W. Pedrycz, Advances in fuzzy clustering and its applications, Wiley-Blackwell (2007) 237-242.

[34] Y. Y. Yang, M. Mahfouf, G. Panoutsos, Probabilistic Characterisation of Model Error Using Gaussian Mixture Model– with Application to Charpy Impact Energy Prediction for Alloy Steel, Control Engineering Practice 20 (2012) 82-92.

[35] G.J. McLachlan, Mixture Models: inference and applications to clustering, Dekker (1988).

[36] G. J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, John Wiley and Sons Inc. (2008), Hoboken, New Jersey.

[37] R. Simon, M. Girolami, A first course in Machine learning, Boca Raton: CRC Press (2012).

[38] J. J. Leader, Numerical Analysis and Scientific Computation, Pearson, United States, 2004.

[39] C. Keary, P. Sheskey, Preliminary report of the discovery of a new pharmaceutical granulation process using foamed aqueous binders, Drug Development and Industrial Pharmacy 30 (2004) 831–845.

[40] T. Schaefer, P. Holm, H. Kristensen, Wet granulation in a laboratory scale high shear mixer, Pharmaceutical Technology 52 (1990) 1147–1153.

[41] G. Reynolds, C. Biggs, A. Salman, M. Hounslow, Non-uniformity of binder distribution in high-shear granulation, Powder Technology 140 (2004) 203-208.

[42] A. Scott, M. Hounslow, T. Instone, Direct evidence of heterogeneity during high-shear granulation, Powder Technology 113 (2000) 205-213.

[43] J. D. Osborne, R. P.J. Sochon, J. J. Cartwright, D. G. Doughty, M. J. Hounslow, A. D. Salman, Binder addition methods and binder distribution in high shear and fluidised bed granulation, Chemical Engineering Research and Design 89 (2011) 553-559.